IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received January 24, 2021, accepted February 9, 2021, date of publication February 12, 2021, date of current version February 24, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3058993

A Novel Framework for Early Pitting Fault Diagnosis of Rotating Machinery Based on Dilated CNN Combined With Spatial Dropout

XUEYI LI[1,2,3], (Graduate Student Member, IEEE), XIANGWEI KONG[1,2,3], (Member, IEEE), ZHENDONG LIU[1], ZHIYONG HU[1,2,3], AND CHENG SHI[4]

[1]School of Mechanical Engineering and Automation, Northeastern University, Shenyang 110819, China
[2]Key Laboratory of Vibration and Control of Aero-Propulsion System, Ministry of Education, Northeastern University, Shenyang 110819, China
[3]Liaoning Province Key Laboratory of Multidisciplinary Design Optimization of Complex Equipment, Northeastern University, Shenyang 110819, China
[4]School of Mechanical Engineering, Beijing Institute of Technology, Beijing 100081, China

Corresponding author: Xiangwei Kong (xwkong@me.neu.edu.cn)

This work was supported in part by the State Ministry of Science and Technology Innovation Fund of China under Grant 2018IM030200, in part by the National Key Research and Development Program of China under Grant SQ2019YFB170107, in part by the National Natural Foundation of China under Grant U1708255, and in part by the China Scholarship Council under Grant 201906080059.

ABSTRACT Pitting corrosion of rotating machinery is one of the most common faults in industrial engineering. The convolutional neural network (CNN) is increasingly applied to the fault diagnosis. However, the conventional CNN method will reduce the feature dimension of the collected signal and cause the loss of information during the pooling process. In this paper, a new method based on dilated CNN combined with spatial dropout (DCSD) is proposed to diagnose the early faults of rotating machinery. By filling the convolution kernel, the DCSD method can increase the receptive field of the CNN without increasing the number of parameters while retaining more features of the raw vibration signal of the rotating machine. To avoid the dropout method eliminates the adjacent elements with a strong correlation, the Spatial Dropout method is adopted to reduce the overfitting problem of deep networks. The early pitting gears experiment was designed to verify the DCSD method in this paper. The raw vibration signal data of 6 different healthy states were collected to verify the effectiveness of the method. The experimental results show that the DCSD method proposed can effectively distinguish the different early gears pitting, and the diagnostic accuracy is better than other popular deep learning methods.

INDEX TERMS Deep learning, dilated convolutional neural network, fault diagnosis, rotating machinery, spatial dropout.

I. INTRODUCTION

Fatigue damage is one of the most common faults of industrial engineering [1]. Because the vibration signal features of early pitting fault of gear are weak, the traditional fault diagnosis method is challenging to find in time. Early fault diagnosis of gear has always been a great challenge. Since the machine has no noticeable symptoms before fatigue damage occurs, it will have disastrous consequences for human society and the ecological environment. Major disasters caused by fatigue fracture of mechanical components are frequent in the world. For example, in the disaster of Sayano-Shushenskaya Dam, 80% of the screw piles had fractured before the accident, but they were overlooked. It was a lack of awareness of early faults that eventually led to the world's worst hydropower station collapse [2]. It is precisely because human society is eager to reduce catastrophic accidents and economic losses that the research on fatigue diagnosis and residual life prediction has long been widely concerned by academia and industry [3]. Despite this, the research on fatigue diagnosis and residual life prediction still cannot meet the requirements of high accuracy and high reliability in industrial applications. The frequent occurrence of disasters indicates that fault diagnosis and residual life prediction are still not mature in practical applications in the industry.

At present, most of the techniques and theories used in fault detection of rotating machinery are confined to machine learning methods. A branch of fault detection is the use

The associate editor coordinating the review of this manuscript and approving it for publication was Dazhong Ma.

of enhanced or estimated noise to improve signal-to-noise ratio (SNR) and extract fault signatures [4]. Yuan *et al.* [5] used two noise estimation techniques and strategies for different SNRs to detect mechanical faults. Hanachi *et al.* [6] applied the Gas Turbine Engine developed to diagnose four kinds of faults that gradually deteriorate to different degrees. The results showed that compared with the previously fault parameter estimation scheme, the diagnostic accuracy was improved by ten times. In terms of denoising, Gan *et al.* [7] proposed a multiple-domain manifold method and singular value decomposition (SVD). The SVD is used to calculate the singular value in the reconstructed space and the improved singular value. Cerrada *et al.* [8] established a robust multistage fault diagnosis system for spur gears by selecting the optimal state parameter set in the time, frequency, and time-frequency domains extracted from vibration signals. Feature selection is still an essential aspect of machine learning-based fault diagnosis to achieve an excellent performance. However, manually selecting features requires not only a lot of prior knowledge but also consumes a lot of time and energy. Conventional methods are labor-intensive because they often rely on handcrafted features that require expertise. As an emerging field in industrial application and an effective solution for fault identification, artificial intelligence technology has attracted more and more attention from academia and industry [9].

The establishment of the Industrial Internet makes real-time online mechanical fault diagnosis become a vital component. Some new detection techniques have also been used in the fault detection in mechanical components, such as acoustic emission [10] and technology based on piezoelectric elements and active Lamb waves [11]. Elforjani *et al.* [12] connected the acoustic emission signal and corresponding natural wear correlation of the slow bearing. The disadvantage of using piezoelectric elements and active Lamb waves for health monitoring of mechanical components is that active Lamb waves can only be generated by providing a voltage power supply. Hence, its application in the health monitoring of complex mechanical structural components is limited [13].

In recent years, the research and application of health monitoring methods based on deep learning [14] are attracting increasing attention in the coming era of big data in the industry [15]. Deep learning models include deep belief network [16], autoencoder [17], restricted Boltzmann machine [18], recurrent neural network (RNN) [19], convolutional neural network (CNN) [20], etc. For different data types, different network models need to be selected to represent the data distribution. For example, for text data, RNN and CNN are usually used. Shao *et al.* [21] proposed a new fault diagnosis framework that used infrared thermal image carried out fault diagnosis for rotor-bearing under variable working conditions. Improved transmission CNN solved the problem of limited available training data in the target domain. He *et al.* [22] proposed an ensemble transfer CNNs driven by multi-channel signals to carry out fault diagnosis of rotating machinery under cross-working conditions. Zhao *et al.* [23] proposed a local feature-based gated recurrent unit (GRU) network. It is a hybrid approach that combines handcrafted functional design with automated learning for machine health monitoring. Long *et al.* [24] combined a competitive swarm optimization algorithm with a local search for intelligent diagnosis of mechanical faults. This method can reliably identify different faults. He *et al.* [25] proposed a method that uses the Short-Time Fourier Transform to preprocess sensor signals. The experimental results show that the method can provide better diagnostic performance even at a relatively low speed. A method proposed by Li *et al.* [26] used the one-dimensional separable convolution with a residual connection network to classify gear faults under mixed conditions. In addition, autoencoders and support vector machines are applied to fault diagnosis of multi-joint industrial robots [27]. By extracting the bottom features layer by layer and gradually aggrading them to abstract features. Although pooling in CNN can increase the receptive field and improve the performance of the network, downsampling will cause the loss of information of the input features. The maximum pooling is to divide the input signal into several sub-area and output the maximum value for each sub-area, and all other values will be discarded. For the early fault diagnosis, due to the weak vibration signal features of the fault, all features except the maximum value are abandoned through maximum pooling, which is not beneficial to the early diagnosis of weak fault signals of rotating machinery. Therefore, pooling operation is not the best method in the early fault diagnosis of rotating machinery.

To solve the problems above, this paper proposes a new method, Dilated CNN combined with spatial dropout (DCSD), for early fault diagnosis of rotating machinery, which uses a dilated CNN instead of convolution and pooling. Compared with ordinary convolution, dilated convolution can increase the receptive field without increasing network parameters or decreasing network dimensions. At the same time, the spatial dropout method is adopted to reduce the overfitting problem of the deep network to avoid the problem that the conventional dropout method eliminates the adjacent features with a strong correlation. Compared with traditional dropout, which randomly sets some elements to zero, spatial dropout randomly sets some areas to zero. This dropout method has been proved to be effective in early fault diagnosis of rotating machinery. The highlights of this paper are summarized as follows: Compared with the conventional one-dimensional CNN, the early fault diagnosis method for rotating machinery proposed increases the receptive field. It maintains excellent diagnostic performance without increasing network parameters or reducing network dimensions.

The rest of this paper is organized as follows. The proposed DCSD method is described in detail in Section II. In Section III, a gearbox test rig was designed to verify the proposed method. In Section IV, the test results are analyzed and discussed, and the effectiveness of the proposed early fault diagnosis method for rotating machinery is validated by

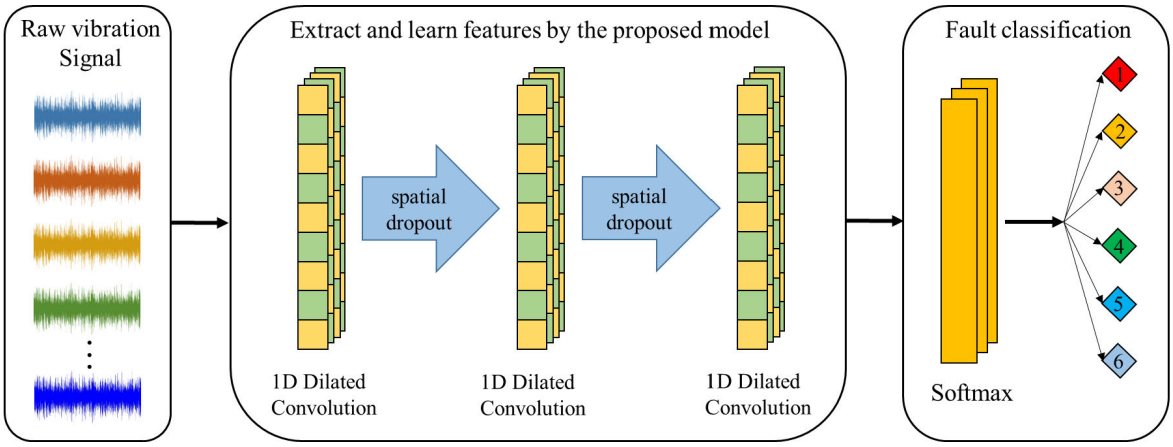


FIGURE 1. Flowchart of the proposed DCSD model.

comparing it with other methods. Finally, the conclusion was drawn in Section V.

II. THE METHOD

The environmental noise of rotating machinery is extensive and complex, and the classical CNN is difficult to meet the industrial needs for early fault diagnosis. In this paper, the dilated CNN and spatial dropout are used for the early fault diagnosis of rotating machinery. Based on the expansion of the convolution kernel, a novel module with a larger receptive field is proposed. Besides, a depth neural network based on spatial dropout and dilated convolution are constructed. The flowchart of the DCSD method for the early fault diagnosis of rotating machinery is shown in Fig. 1. The raw vibration signal of rotating machinery is taken as input. The feature extraction module is composed of 1D dilated CNNs and 1D spatial dropouts. The collected features are input into the classification module Softmax for early fault diagnosis of rotating machinery.

A. DILATED CONVOLUTIONAL NEURAL NETWORKS

In the early fault diagnosis of rotating machinery, the more information extracted from the inputs, the more likely the raw vibration signals will be correctly classified. The receptive field is feeling scope size relative to the original input signal. The size of the receptive field represents how many features the extracted. The larger the receptive field is, the more information it contains. Therefore, more features can be obtained by increasing the receptive field of neurons in CNN. If the receptive field contains the early fault diagnosis characteristics, it is easier to identify the early fault correctly.

However, if only consider increasing the receptive field and performing excessive pooling operations, the signal will lose too many features. In the early fault diagnosis of rotating machinery, CNN will pass through the pooling layer many times. Each pooling layer is downsampling, and the receptive field of neuron nodes will be increased by reducing the size of the original vibration signal [28]. Assuming that the pooling step of each pooling layer is 2, each time through the pooling layer, the output feature becomes 1/2 of the input feature. The input signal is continuously pooled five times, and the final output becomes 1/32 of the input signal. In this process, the output features become very sparse, which is not conducive to the early fault diagnosis of rotating machinery. Therefore, a method is needed to make the original vibration signal denser and the receptive field larger. The dilated convolution just meets these two requirements. It can have a relatively sizeable receptive field without downsampling.

The dilated convolution kernel is to inject holes into the standard convolution kernel to increase the receptive field. The weight of the newly added convolution point is set as 0, and dilated convolution kernels are obtained. Fig. 2 shows the way to fill the convolution kernel. The size of the original convolution kernel is 3×3. When the dilation rate is 2, a hole is inserted between the elements of the convolution kernel to obtain a dilated convolution kernel with a size of 5×5.

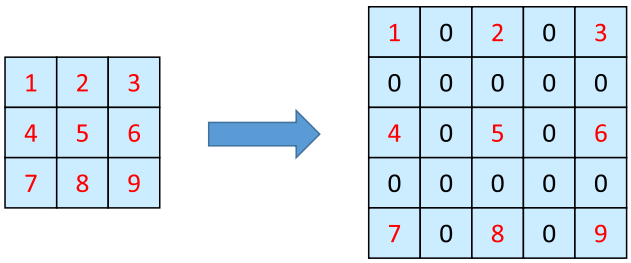


FIGURE 2. Filling principle of the dilated convolution kernel (on the left is the conventional convolution kernel, right is the Dilated convolution kernel).

For a convolution kernel with a size of $k \times k$, assuming that the size of the convolution kernel after inserting the hole is $k_d \times k_d$, r on behalf of dilation rate, then

$$k_d = k + (k-1)(r-1) \tag{1}$$

A schematic diagram of dilated convolution quoted from Princeton University Fisher Yu *et al.* [29] was shown in Fig 3.

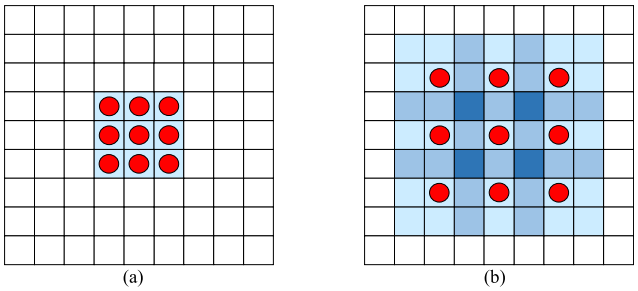


FIGURE 3. The receptive field of convolution kernel (on the left is the conventional CNN operation, right is the dilated CNN operation).

The red dot in the figure represents the coordinates covered by the convolution kernel, and the blue part represents the convolution operation. In conventional CNN, the parameters in a convolution kernel will be multiplied and added with the values on the feature map in a continuous range, as shown in the left figure. Fig. 3(a) is a standard 3×3 convolution operation. That is, the dilation rate is 1, and the receptive field is a 3×3 area. Fig. 3(b) is a 3×3 conventional convolution kernel connected in series 3×3 dilated convolution kernel, each red dot will cover a 3×3 area, so the overall area is nine 3×3 convolution kernels with a step size 2, forming a 7×7 area. At this time, the convolution kernel can be regarded as a kernel size of 7×7. However, there are only nine useful convolutions. It can be seen that the receptive field expands to 7×7 when the calculated amount unchanged. The light blue blocks in the figure have only undergone one convolution operation, and the darker blue blocks have undergone multiple convolution operations. By analogy, when the distance between adjacent weights is 4, it is equivalent to nine 7×7 areas spread out by step of 4 to form a 15×15 area. Compared with the number of parameters of the standard convolution, the number of parameters of the dilated convolution has not increased. Still, the receptive field is significantly larger than that of the standard convolution

Fig. 4 shows a schematic diagram of an ordinary convolution operation and a dilated convolution operation. Take an input feature map with a size of 7×7 and a convolution kernel with 3×3 as an example. Fig. 4 (a) shows an ordinary CNN operation. Fig. 4 (b) shows a dilated CNN operation with a dilated rate is 2. Intuitively, the latter is the same thing as a 5×5 convolution kernel. The larger the convolution kernel, the larger the receptive field of the network, so the dilated convolution can use a smaller convolution kernel to obtain a larger receptive field.

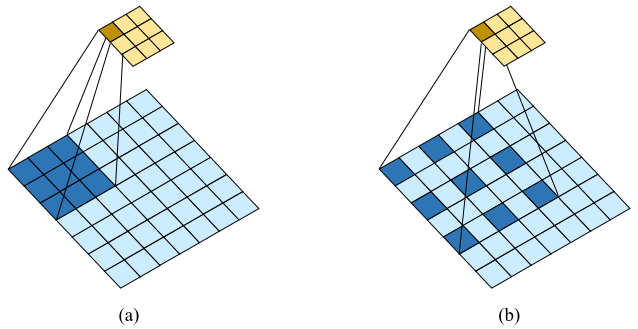


FIGURE 4. Diagram of conventional CNN and dilated CNN operation.

In the basic CNN structure, when the convolution step is fixed at 1, the receptive field recursion formula of ordinary convolution is shown as follows:

$$F_n = F_{n-1} + K_n - 1 \tag{2}$$

where F_n is the length of the receptive field of the back layer, F_{n-1} is that of the front layer, and K_n is the length of the convolution kernel between the front and rear layers. For convolution with space, its receptive field recurrence formula is as follows:

$$F_n = F_{n-1} + D_n \times (K_n - 1) \tag{3}$$

where D_n represents the dilation rate between the convolution of the front and the back layers, and the other terms are the same as (1). According to the above method, the most direct ways to expand the receptive field include increasing the depth of the network, increasing the step size parameter of the convolution kernel, and increasing the size of the convolution kernel.

The size of the output feature map of the dilated convolution is:

$$o = \frac{W + 2p - k - (k-1) \times (r-1)}{s} \tag{4}$$

where k represents the size of the original convolution kernel, r represents the expansion rate of the dilated convolution relative to the original convolution, o represents the size of the feature output by the dilated convolution, W represents the size of the feature input, p represents the data padding size, and s represents the slide step of the convolution.

It can be seen from the above process that the dilated convolution can increase the receptive field by injecting holes in the convolution kernel and ensure that the calculation amount of the network does not increase significantly. This method overcomes the problems of feature detail information, spatial information loss, and small target information fault caused by sampling.

For a one-dimensional input signal $x[i]$, a filter $w[k]$ with length k and expansion coefficient l is used for dilated convolution operation, which can be expressed as (5).

$$y[i] = \sum_{k=1}^{k} x[i + s \times k] w[k] \tag{5}$$

Conventional CNN can be regarded as a dilated convolution with a dilation rate is 1. When the dilation rate is greater than 1, the effect of increasing the receptive field can be achieved, as shown in Fig. 5. In Fig. 5, (a) is the conventional CNN with $k = 3$, and the step size is 2. At this time, the receptive field size of the latter layer of the network is 3; (b) is the conventional CNN with $k = 3$, and the step size is 1; (c) is the dilated CNN with $k = 3$, the step size is 1, and the dilation rate is 2. Although the size of the convolution kernel

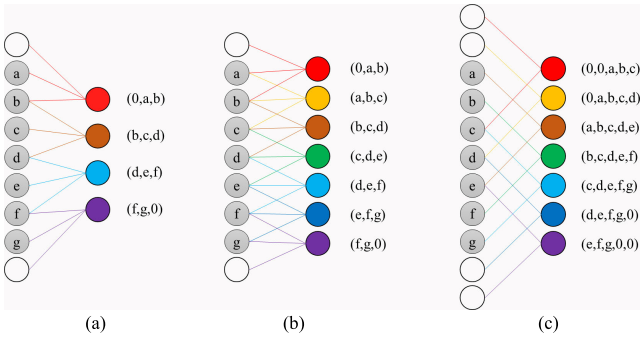


FIGURE 5. 1D dilated convolution receptive field.

in (c) is 3, every other point is sampled, and the receptive field size is 5. The content in the parentheses after the output node is the name of the input node, and the hollow circle represents the node with zero paddings.

B. SPATIAL DROPOUT

The dropout operation randomly zeroes some elements and performs a scale transformation for the non-zero parts [30]. The magnitude of the scale conversion is related to the drop rate. The specific process can be understood as follows: After the zeroing operation, the zero part can be considered to be discarded. Therefore, the model is forced to use the remaining information to fit the target. However, each dropout is random, so the model is forced to learn with a relatively small number of features, each time the learned features are different. Each feature should contribute to the model's prediction, rather than focusing on partial features, leading to overfitting.

Spatial dropout is a novel dropout method proposed by Tompson *et al.* [31]. Spatial dropout is similar to dropout, but it disconnects the entire feature map instead of a single feature. If there is a strong correlation between adjacent features, ordinary dropout cannot normalize its output. Spatial dropout can help improve independence between features. When using dropout technology, it can be found that normal dropout randomly and independently sets some elements to zero, while spatial dropout randomly sets all to zero for a specific dimension, as shown in Fig. 6:

C. RELU ACTIVATION FUNCTIONS

The Rectified Linear Unit (ReLU) activation function was first proposed by Hinton *et al.* [32]. It is a more straightforward function than the Sigmoid function [33] and can effectively avoid the problem of gradient disappearance. In this paper, the ReLU activation function was used as the feature map of the early pitting signal of rotating machinery. Compared with the Sigmoid function and Tanh function [34], ReLU has a small amount of calculation, high computational efficiency, and can alleviate the problem of gradient disappearance. At the same time, ReLU has the characteristics of sparseness and alleviates the problem of overfitting [35]. Therefore, it has gradually replaced the Sigmoid activation function and has become the most widely used activation function in the early fault diagnosis tasks of rotating machinery. The mathematical expression of the ReLU function and its derivative function expression is:

$$f(u) = \max(0, u) \tag{6}$$

$$f'(u) = \begin{cases} 0, & u < 0 \\ 1, & u > 0 \end{cases} \tag{7}$$

The ReLU function and its derivative function image are shown in Fig. 7. In Fig. 7(a), the abscissa u represents the output of the previous network, and the ordinate f(u) denotes the output of the ReLU network. When the input u >0, the output of the activation function is equal to the input; otherwise, the output value is 0. The image of the derivative of the ReLU function is shown in Fig. 7(b). When the input is negative, the derivative of the ReLU function is 0; when the input is positive, the derivative of the ReLU function is 1, and no gradient disappears phenomenon.

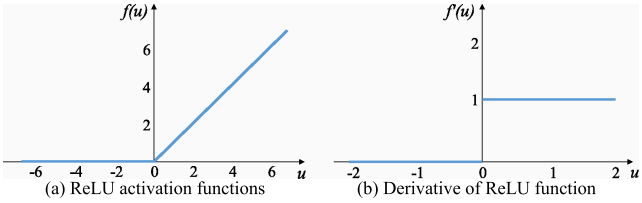


FIGURE 7. ReLU function and its derivative image.

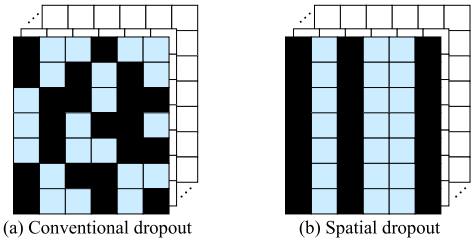


FIGURE 6. Schematic diagram of conventional dropout and spatial dropout.

III. GEAR TEST EXPERIMENTS

To verify the DCSD method proposed, a gear early pitting experiment was designed. As shown in Fig. 8, the experimental equipment is mainly composed of two 45kW motors, a gearbox, and a cooling device. The raw vibration data is transferred to the notebook computer through the data acquisition card through the accelerometer installed on the gearbox. The vibration signal sampling rate is 10240Hz; the sampling time is 3s; the data were collected independently 5 times. The speed is 1800rpm, and the load is 50Nm.

The experiments used six different healthy state gears. As shown in Fig. 9, the gear tooth of the healthy state 1 was not pitting; the intermediate gear tooth in the healthy state 2 have about 10% pitting; The intermediate gear tooth in the health state 3 have about 30% pitting; the intermediate gear tooth in the health state 4 has about 50% pitting. All the gear

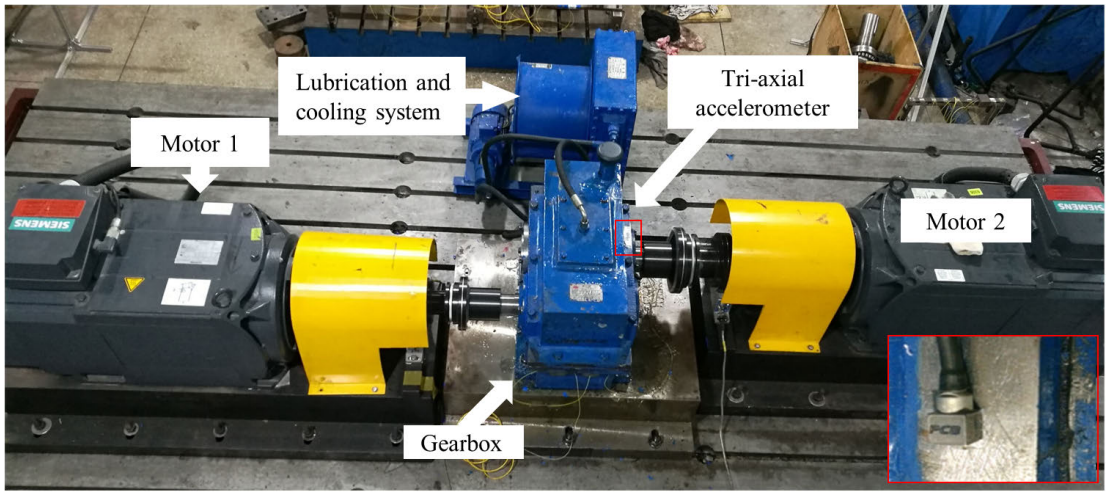


FIGURE 8. Photo of the test bench.

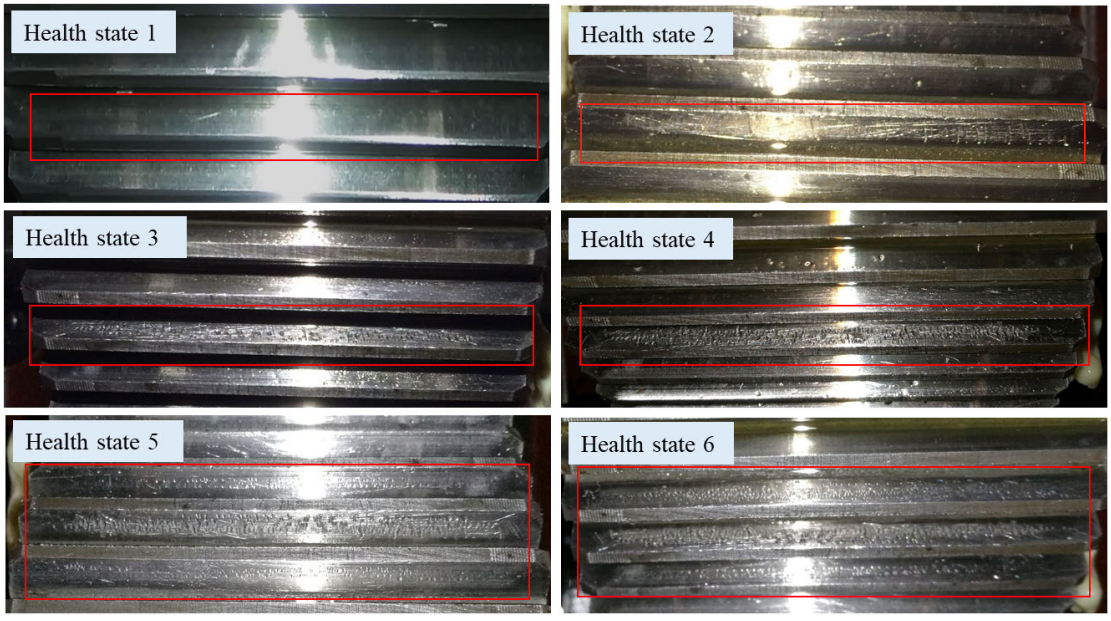


FIGURE 9. Photos of 6 kinds of different health state gears.

teeth were normal except the intermediate tooth in the healthy states 2, 3, and 4. The intermediate gear tooth in the health state 5 has about 50% pitting, and two adjacent gear teeth have 10% pitting, respectively. The intermediate gear teeth in the healthy state 6 gear has about 50% pitting; the adjacent upper gear tooth has 30% pitting, and the adjacent lower gear tooth has 10% pitting.

The raw gears vibration signals in 6 kinds of health states are shown in Fig. 10. The gears in each health state continuously extract 10,000 points for drawing. The abscissa is the number of the extracted points, and the ordinate is the amplitude. It can be seen from Fig. 10 that the amplitude range of the 6 kinds of health states is between -4 m/s2 and 4 m/s2, and the average value is around 0. The amplitude of gears in health state 1 is slightly larger than that of gears in other health states, and the amplitude of gears in health state 6 is slightly smaller than that of gears in other health states. It is difficult to distinguish the image of the raw vibration signal in different health states with the naked eye. Conventional machine learning methods are also difficult to distinguish such close vibration signals effectively.

The frequency histogram of the amplitude of the raw gears vibration signal in six different health states is shown in Fig. 11, in which 10,000 points were extracted for drawing. From the figure, all of the frequency histograms are following the normal distribution. The largest number of frequency is around 600, and the smallest frequency part is around ± 3.5. The raw vibration signals of different health states in this experiment are very close, and the signal is not well partitioned.

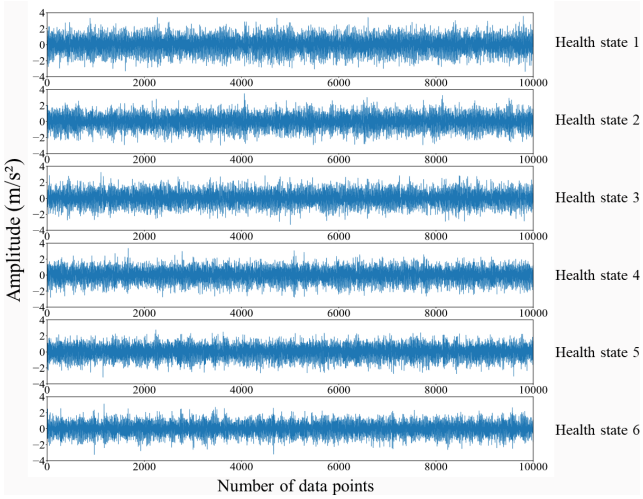


FIGURE 10. Raw vibration signal of gears in 6 kinds of health states.

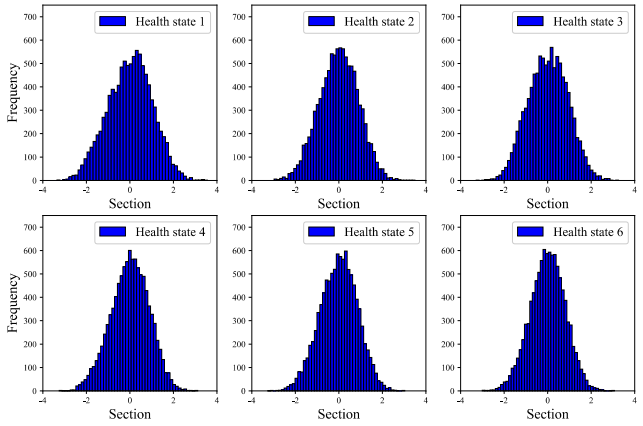


FIGURE 11. The frequency histogram of the amplitude of the raw vibration signal in 6 health states.

In order to adapt to the neural network sensitive interval, the raw vibration signal was contracted to between 0 and 1. The vibration signal box plot of the 5 data sets in 6 different health states is shown in Fig. 12. It can be seen from the figure that all the data set median concentrate between 0.4 and 0.5, which is relatively stable. The upper quartile is around 0.5, and the lower quartile is around 0.35. The positions of the farthest points of health state 1, 2, and 6 fluctuate considerably, ranging from 0.7-0.8 and 0.1-0.2, respectively. The positions of the farthest points of the gears of other health states are relatively stable, at nearby 0.25 and 0.7, respectively. The abnormal value in health state 1 is relatively more, and that in the other five health states are controlled in single digits. In general, the data of this verification experiment is relatively clean. Although there are outliers, the number of outliers accounts for a small proportion. The data can be used to validate the DCSD method presented in this paper.

In this validation experiment, six kinds of gear early pitting faults were accurately distinguished. Five data sets of gear vibration signals for each healthy state, including four sets, are used for training, and one set for testing. The network uses three dilated convolutional layers and two spatial dropout layers stacked. Dilated convolution can significantly increase the receptive field and retain the detailed information of the raw vibration signal. Finally, a dense layer and a softmax layer are connected, and the neural network output is normalized to the probability distribution of (0, 1) through the softmax classifier. Add Batch Normalization [36] layers after each convolutional layer to normalize the output data to a standard normal distribution. The hyperparameters of the network are as follows: For the dilated convolution, the channel is 32, 64, and 128; kernel size is 9; strides is 1; the activation function is ReLU; dilation rate is 2; bias initializer is random normal, and the standard deviation is 0.01. spatial dropout takes 0.5. The optimization method adopts SGD. The initial learning rate is set to 0.4. Network training was done through an Alienware laptop, the software used Python, and the Tensorflow framework was used for training. The main configuration of the computer is as follows: Intel Core i7-9750H, 32 GB RAM, NVMe 512GB solid-state drive, and Nvidia GeForce GTX 1660 Ti.

IV. RESULTS AND DISCUSSIONS

The accuracy curves of the training set and the validation set of the method proposed in this paper are shown in Fig. 13. It can be seen from the figure that the accuracy curve of the training set rises faster in the initial stage. The accuracy curve of the training set is higher than that of the validation set. The two curves gradually merged from about the 40th epoch. They overlapped after about the 60th epoch. It can be seen from the results that the method proposed in this paper can effectively judge the early pitting faults of gears.

The loss curves of the training set and validation set of the method proposed in this paper are shown in Fig. 14. From the figure, it can be seen that the loss curve of the training set decreases faster in the initial stage, and the loss curve of the training set is lower than that of the validation set. For the loss curve, the two curves gradually merge from about the 40th epoch and overlap after about the 60th epoch. It can be seen from the results that the method proposed in this paper has a good effect on the diagnosis of early gear pitting fault diagnosis.

The confusion matrix of the DCSD method proposed in this paper for six different health states is shown in Fig. 15. It can be seen from the figure that for the early gear pitting in health states 3 and 6, the method proposed to predict the accuracy rate reached 100%. And 1% of early gear pitting with health states 1 and 5 were misjudged as health states 2 and 4, respectively. 2% of the early gear pitting with health states of 2 and 4 were misjudged as health states of 3 and 6, respectively. From Fig. 10-12, the early pitting signals of different healthy states are very similar, and the influence of environmental noise may cause the neural network to misjudge. But the overall classification effect for early gear pitting is still excellent.

It is comparing to the DCSD method proposed in this paper with other classic deep learning methods. Performing

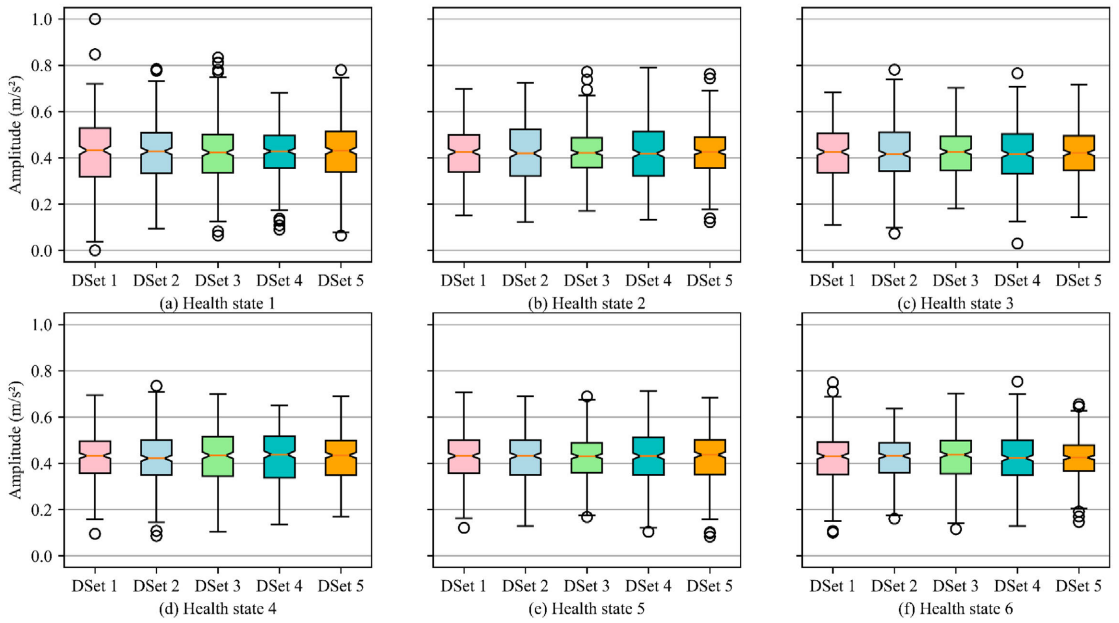


FIGURE 12. The raw vibration signal box plot of the 5 data sets in 6 different health states.

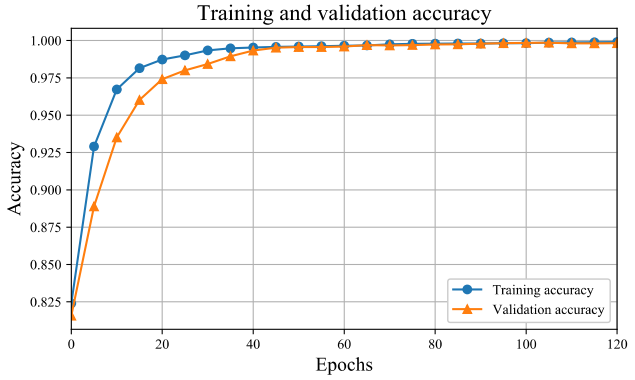


FIGURE 13. The accuracy curves of the training set and validation set of the method proposed.

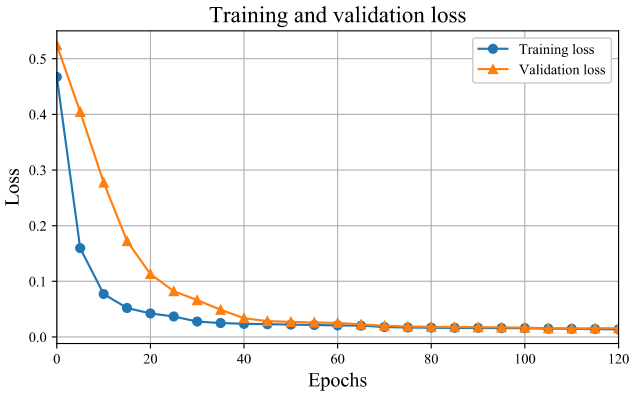


FIGURE 14. The loss curves of the training set and validation set of the method proposed.

on the same gear early fault diagnosis data, and the result is shown in Fig. 16. The parameters for the methods shown in Fig. 16 are listed in Table 1. The DCSD method proposed in this paper has the best effect on the early gears pitting fault diagnosis, and the diagnosis accuracy rate exceeds 99%.

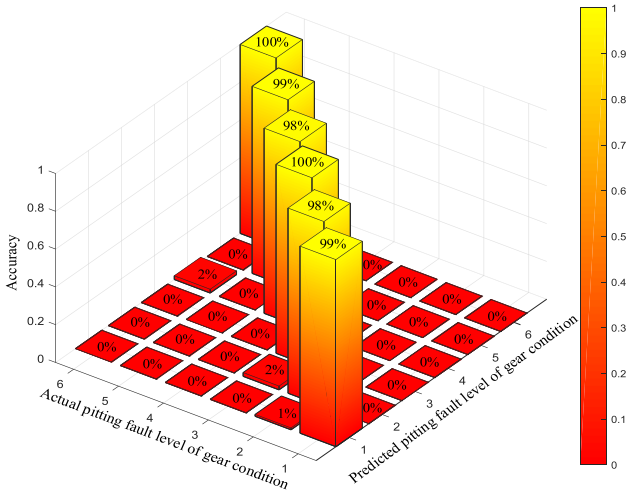


FIGURE 15. The confusion matrix of six gears with different health states.

The stability of the DCSD method is excellent. The standard deviation of the method proposed is within ±0.5. Although the other methods have achieved relatively good results, the accuracy rate was all above 80%, but there is still a gap with the DCSD method. It can be seen from the results that the method proposed in this paper is very effective for early gear pitting fault diagnosis.

For each type of gear pitting healthy states, 100 points are selected for t-SNE visualization. The results of the visualization are shown in Fig. 17. The six different gear health states have relatively clear boundaries on the three-dimensional (3D) coordinate map. Although the boundary between a few gears is not particularly bright, most of the raw vibration signals can be effectively classified. The method proposed in this paper is useful for the early pitting fault diagnosis of rotating machinery.

TABLE 1. The parameters of the DCSD method and that of the other methods.

	Channel	Kernel size	Dilation rate	Dropout	Cell size	Learning rate
Proposed method	32, 64, 128	9	2	0.5	—	0.4
Traditional CNN with spatial dropout	32, 64, 128	9	2	0.5	—	0.4
Dilated CNN with traditional dropout	32, 64, 128	9	2	0.5	—	0.4
Gate recurrent unit	—	—	—	—	32,32,16	0.1
Long short term memory	—	—	—	—	32,32,16	0.1
Traditional CNN only	32, 64, 128	9	2	—	—	0.4
Bidirectional long short term memory	—	—	—	—	32,32,16	0.1

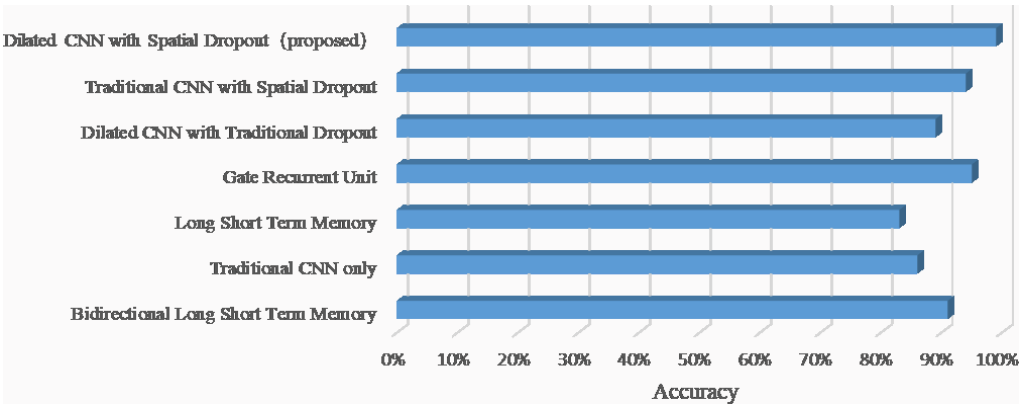


FIGURE 16. Comparison of the accuracy of the DCSD method with other methods.

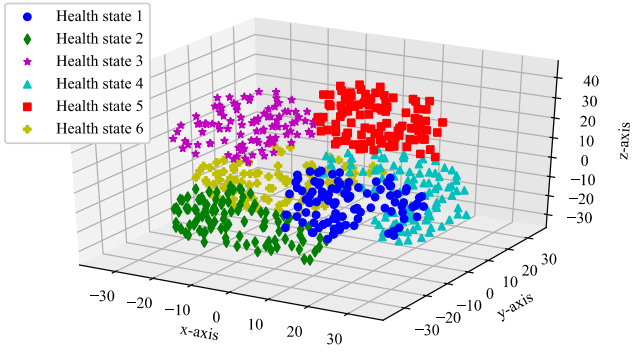


FIGURE 17. 3D visualization of 6 different pitting health states.

V. CONCLUSION

This paper proposes a novel deep learning method based on the combination of dilated convolution and spatial dropout for early fault diagnosis of rotating machinery. The verify experiment proves the effectiveness of the DCSD method for six kinds of early gears pitting fault. The early fault diagnosis method for rotating machinery proposed maintains excellent diagnostic performance without increasing network parameters or reducing network dimensions. The superiority of this method is verified through comparative research. As a result, the DCSD method can be applied to the early fault diagnosis of rotating machinery effectively. The author will conduct further research in the future.

ACKNOWLEDGMENT

The authors acknowledge Dr. Y. Qu, who designed the experiments and performed the data collection.

REFERENCES

[1] J. Sun, C. Yan, and J. Wen, "Intelligent bearing fault diagnosis method combining compressed data acquisition and deep learning," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 1, pp. 185–195, Jan. 2018.

[2] J. Schijve, "Fatigue of structures and materials in the 20th century and the state of the art," *Int. J. Fatigue*, vol. 25, no. 8, pp. 679–702, Aug. 2003.

[3] B. Samanta, "Gear fault detection using artificial neural networks and support vector machines with genetic algorithms," *Mech. Syst. Signal Process.*, vol. 18, no. 3, pp. 625–644, 2004.

[4] H. Heidari Bafroui and A. Ohadi, "Application of wavelet energy and Shannon entropy for feature extraction in gearbox fault detection under varying speed conditions," *Neurocomputing*, vol. 133, pp. 437–445, Jun. 2014.

[5] J. Yuan, F. Ji, Y. Gao, J. Zhu, C. Wei, and Y. Zhou, "Integrated ensemble noise-reconstructed empirical mode decomposition for mechanical fault detection," *Mech. Syst. Signal Process.*, vol. 104, pp. 323–346, May 2018.

[6] H. Hanachi, J. Liu, I. Y. Kim, and C. K. Mechefske, "Hybrid sequential fault estimation for multi-mode diagnosis of gas turbine engines," *Mech. Syst. Signal Process.*, vol. 115, pp. 255–268, Jan. 2019.

[7] M. Gan, C. Wang, and C. Zhu, "Multiple-domain manifold for feature extraction in machinery fault diagnosis," *Measurement*, vol. 75, pp. 76–91, Nov. 2015.

[8] M. Cerrada, G. Zurita, D. Cabrera, R.-V. Sánchez, M. Artés, and C. Li, "Fault diagnosis in spur gears based on genetic algorithm and random forest," *Mech. Syst. Signal Process.*, vols. 70–71, pp. 87–103, Mar. 2016.

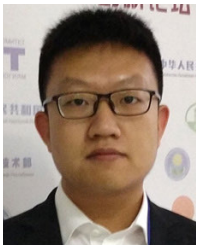
[9] R. Liu, B. Yang, E. Zio, and X. Chen, "Artificial intelligence for fault diagnosis of rotating machinery: A review," *Mech. Syst. Signal Process.*, vol. 108, pp. 33–47, Aug. 2018.

[10] X. Li, J. Li, D. He, and Y. Qu, "Gear pitting fault diagnosis using raw acoustic emission signal based on deep learning," *Eksploatacja i Niezawodność*, vol. 21, no. 3, p. 403, 2019.

[11] W. Yang and P. Gao, "Lamb wave-minimum sampling variance particle filter-based fatigue crack prognosis," *Sensors*, vol. 19, no. 5, p. 1070, Mar. 2019.

[12] M. Elforjani and S. Shanbr, "Prognosis of bearing acoustic emission signals using supervised machine learning," *IEEE Trans. Ind. Electron.*, vol. 65, no. 7, pp. 5864–5871, Jul. 2018.

[13] J. C. Dodson and D. J. Inman, "Thermal sensitivity of Lamb waves for structural health monitoring applications," *Ultrasonics*, vol. 53, no. 3, pp. 677–685, Mar. 2013.
[14] X. Ding and Q. He, "Energy-fluctuated multiscale feature learning with deep ConvNet for intelligent spindle bearing fault diagnosis," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 8, pp. 1926–1935, Aug. 2017.
[15] X. Li, J. Li, Y. Qu, and D. He, "Gear pitting fault diagnosis using integrated CNN and GRU network with both vibration and acoustic emission signals," *Appl. Sci.*, vol. 9, no. 4, p. 768, Feb. 2019.
[16] Z. Chen and W. Li, "Multisensor feature fusion for bearing fault diagnosis using sparse autoencoder and deep belief network," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 7, pp. 1693–1702, Jul. 2017.
[17] X. Li, J. Li, Y. Qu, and D. He, "Semi-supervised gear fault diagnosis using raw vibration signal based on deep learning," *Chin. J. Aeronaut.*, vol. 33, no. 2, pp. 418–426, Feb. 2020.
[18] C. Li, R.-V. Sánchez, G. Zurita, M. Cerrada, and D. Cabrera, "Fault diagnosis for rotating machinery using vibration measurement deep statistical feature learning," *Sensors*, vol. 16, no. 6, p. 895, Jun. 2016.
[19] H. Liu, J. Zhou, Y. Zheng, W. Jiang, and Y. Zhang, "Fault diagnosis of rolling bearings with recurrent neural network-based autoencoders," *ISA Trans.*, vol. 77, pp. 167–178, Jun. 2018.
[20] O. Janssens, V. Slavkovikj, B. Vervisch, K. Stockman, M. Loccufier, S. Verstockt, R. Van de Walle, and S. Van Hoecke, "Convolutional neural network based fault detection for rotating machinery," *J. Sound Vib.*, vol. 377, pp. 331–345, Sep. 2016.
[21] H. Shao, M. Xia, G. Han, Y. Zhang, and J. Wan, "Intelligent fault diagnosis of rotor-bearing system under varying working conditions with modified transfer CNN and thermal images," *IEEE Trans. Ind. Informat.*, early access, Jun. 30, 2020, doi: 10.1109/TII.2020.3005965.
[22] Z. He, H. Shao, X. Zhong, and X. Zhao, "Ensemble transfer CNNs driven by multi-channel signals for fault diagnosis of rotating machinery cross working conditions," *Knowl.-Based Syst.*, vol. 207, Nov. 2020, Art. no. 106396.
[23] R. Zhao, D. Wang, R. Yan, K. Mao, F. Shen, and J. Wang, "Machine health monitoring using local feature-based gated recurrent unit networks," *IEEE Trans. Ind. Electron.*, vol. 65, no. 2, pp. 1539–1548, Feb. 2018, doi: 10.1109/Tie.2017.2733438.
[24] J. Long, S. Zhang, and C. Li, "Evolving deep echo state networks for intelligent fault diagnosis," *IEEE Trans. Ind. Informat.*, vol. 16, no. 7, pp. 4928–4937, Jul. 2020.
[25] M. He and D. He, "Deep learning based approach for bearing fault diagnosis," *IEEE Trans. Ind. Appl.*, vol. 53, no. 3, pp. 3057–3065, May 2017.
[26] X. Li, J. Li, C. Zhao, Y. Qu, and D. He, "Gear pitting fault diagnosis with mixed operating conditions based on adaptive 1D separable convolution with residual connection," *Mech. Syst. Signal Process.*, vol. 142, Aug. 2020, Art. no. 106740.
[27] J. Long, J. Mou, L. Zhang, S. Zhang, and C. Li, "Attitude data-based deep hybrid learning architecture for intelligent fault diagnosis of multi-joint industrial robots," *J. Manuf. Syst.*, to be published, doi: 10.1016/j.jmsy.2020.08.010.
[28] M. Xia, T. Li, L. Xu, L. Liu, and C. W. de Silva, "Fault diagnosis for rotating machinery using multiple sensors and convolutional neural networks," *IEEE/ASME Trans. Mechatronics*, vol. 23, no. 1, pp. 101–110, Feb. 2018.
[29] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," 2015, *arXiv:1511.07122*. [Online]. Available: http://arxiv.org/abs/1511.07122
[30] A. Achille and S. Soatto, "Information dropout: Learning optimal representations through noisy computation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 12, pp. 2897–2905, Dec. 2018.
[31] J. Tompson, R. Goroshin, A. Jain, Y. LeCun, and C. Bregler, "Efficient object localization using convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 648–656.
[32] G. E. Dahl, T. N. Sainath, and G. E. Hinton, "Improving deep neural networks for LVCSR using rectified linear units and dropout," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2013, pp. 8609–8613.
[33] M. Tanaka, "Weighted sigmoid gate unit for an activation function of deep neural network," *Pattern Recognit. Lett.*, vol. 135, pp. 354–359, Jul. 2020.
[34] F. Godin, J. Degrave, J. Dambre, and W. De Neve, "Dual rectified linear units (DReLUs): A replacement for tanh activation functions in quasi-recurrent neural networks," *Pattern Recognit. Lett.*, vol. 116, pp. 8–14, Dec. 2018.
[35] G. Lin and W. Shen, "Research on convolutional neural network based on improved relu piecewise activation function," *Procedia Comput. Sci.*, vol. 131, pp. 977–984, 2018.
[36] S. Santurkar, D. Tsipras, A. Ilyas, and A. Madry, "How does batch normalization help optimization?" in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, pp. 2483–2493.

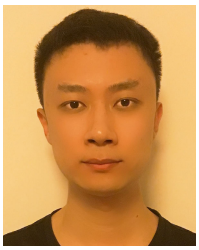


XUEYI LI (Graduate Student Member, IEEE) was born in Harbin, China. He received the master's degree in mechanical engineering from Northeastern University, Shenyang, China, in 2017, where he is currently pursuing the Ph.D. degree with the School of Mechanical Engineering and Automation.

From 2019 to 2020, he was a Visiting Student with the University of Illinois at Chicago, Chicago, USA. His research interest includes intelligent fault diagnostics of rotating machinery.

XIANGWEI KONG (Member, IEEE) was born in Anshan, China. He received the B.S. degree from the Wuhan University of Technology, China, in 1992, and the M.S. and Ph.D. degrees from Northeastern University, Shenyang, China, in 1999 and 2003, respectively.

He is currently a Professor with the School of Mechanical Engineering and Automation, Northeastern University. His research interest includes complex equipment fault diagnosis.



ZHENDONG LIU was born in Liaoyang, China. He received the B.S. degree from the Inner Mongolia University of Technology, Baotou, China, in 2012, and the master's degree in mechanical engineering from Northeastern University, Shenyang, China, in 2018. He is currently pursuing the Ph.D. degree with the School of Mechanical Engineering and Automation, Northeastern University, China.

His research interest includes model based and data-driven method life prediction and fault diagnosis.

ZHIYONG HU was born in Hanzhong, China. He received the B.S. degree from the School of Business Administration, Northeastem University, Shenyang, China, in 2000, and the M.S. and Ph.D. degrees from the School of Mechanical Engineering and Automation, Northeastern University, in 2005 and 2013, respectively.

He is currently a Lecturer with Northeastern University. His research interests include system reliability, intelligent diagnosis, and maintenance.

CHENG SHI was born in Yichun, China. He received the bachelor's degree in vehicle engineering from the Heilongjiang Institute of Technology, Harbin, China, in 2013, and the master's degree from Guangxi University, Nanning, China, in 2017. He is currently pursuing the Ph.D. degree with the School of Mechanical Engineering, Beijing Institute of Technology, China.

His research interest includes modeling and fault diagnosis in the rotary machine.